

Educational Data Mining to support Distance Learning students with difficulties in the Portuguese Grammar

Fabiana M. Soares¹, Crystiano J. R. Machado¹, Diógenes Diniz¹, Alexandre M.A. Maciel¹, Rodrigo L. Rodrigues²

¹Escola Politécnica – Universidade de Pernambuco (UPE) – R. Benfica, 455 – Madalena – CEP: 50.720-001 – Recife/PE

²Centro de Informática - Universidade Federal de Pernambuco (UFPE) – Av. Jornalista Aníbal Fernandes – Cidade Universitária – CEP: 50.740-560 – Recife/PE

{fbms, cjrm, dbod, amam}@ecom.poli.br, rlr@ded.ufrpe.br

***Abstract.** This work proposes the integration of a grammar checker and some EDM techniques in order to cluster students in a LMS according to their grammar difficulties in the Brazilian Portuguese Grammar. It used discussion forums' posts and the results showed that it is possible to identify students' main grammar difficulties and also to group them according to some specific ones.*

1. Introduction

The last census for higher education published in 2014 indicates an increase of newcomers in colleges and universities in Brazil. The number in the Distance Learning (DL) modality has increased 32.9% whereas the On-campus Courses have showed a number only 10% higher [INEP 2014]. Students in both modalities, but especially the ones who take distance courses, use mostly in Brazil the open source LMS (Learning Management System) Moodle (67.3%, according to the census EaD.br) [ABED 2014]. Moodle offers lots of different resources and tools to be used for teaching and learning. One of the most used resources is the Discussion Forum. With it, students and teachers can, for instance, exchange ideas about different subjects and take out doubts through written messages.

Unfortunately, a good number of grammatical difficulties can be observed in these written dialogues. Problems related to students' writings in Portuguese are not novelties among undergraduates. Portuguese is considered a complicated language with a difficult grammar. There are studies in Brazil which shows that students at universities have difficulties in producing pieces of writing [Amaral 2011, Motta 2010]. Add to it the low-quality education offered in the first years of school in Brazil. Consequently, this fact has helped students to accumulate grammar and spelling problems throughout their academic lives. As a result, these problems reach universities and can be observed by teachers from different areas.

Although these mistakes should not be found in the writings of undergraduates, they are there. Teachers often have difficulties to correct them in the discussion forums, not only because of the volume of posts, but also because it is a public area, where everyone from the course can read them. Lots of works have been done regarding discussion forums lately. In Azevedo, Behar and Reategui (2011), for instance, qualitative analyses of textual contributions were made using graphs in order to identify the relevance of a student's answer to a posted question. The idea was to assess students

and bring them back to the topic under discussion. In Romero *et al.* (2013), classification and clustering algorithms were used in order to predict students' final performance according to their participation in online discussion forums. Despite these and others initiatives, there are no works that specifically treat Portuguese grammatical mistakes among undergraduates.

There are different ways to analyze grammatical mistakes. Orthography and grammar checkers are types of applications that use Natural Language Processing (NLP), an area connected to the field of Artificial Intelligence which is concerned with the interactions between computers and natural languages. Considering the educational field, it is possible to use statistical, machine learning and data mining algorithms in order to find knowledge over different types of educational data. So, Educational Data Mining (EDM) explores the data contained in LMS in order to better understand students and their interactions [Romero and Ventura 2010].

Thus, the aim of this work is to propose a workflow to check grammar correctness through Natural Language Processing (NLP) and data mining techniques in order to support DL students who use the LMS Moodle. To do so, the paper is divided as follows: the Related Technologies Study section describes EDM, clustering techniques and the grammar checker used in the project; the entire workflow, which integrates open source applications, is presented and each stage is detailed in section 3; the experiments, from the data extraction to the use of clustering algorithms and the analyses of the results are described in section 4; and in section 5, some suggestions for future works are delivered, as well as the conclusion of the work.

2. Related Technologies Study

This work relates subjects in the educational and computational areas. So, it will describe in the section 2.1 Educational Data Mining (EDM) and intelligent algorithms for clustering and in 2.2 the grammar checker CoGrOO.

2.1 EDM and intelligent algorithms for clustering

Educational Data Mining (EDM) is a field that exploits educational data in distance learning platforms. It discovers new knowledge based on students', teachers', tutors' and gestors' usage of the system in order to improve the quality of the education offered through online courses [Romero and Ventura 2010].

In their review of the state of the art about EDM, Romero and Ventura (2010) lists eleven educational tasks that make use of data mining techniques. Among these issues there is at least a couple which relates to the objective of this work. In the task 'Providing Feedback for Supporting Instructors' they mention that different types of text and data mining techniques, such as decision trees, text segmentation, word segmentation, POS tagging, have been used to analyze, summarize, evaluate, and identify patterns of interaction in texts posted in LMS [Lin, Hsieh and Chuang 2009].

Table 1. Educational Tasks X Examples of Data Mining Techniques [Romero and Ventura 2010]

Educational Tasks	DM Techniques Used
Analysis and Visualization of data	Statistics and visualization information
Providing feedback for supporting instructors	Association rules, clustering, classification, sequential pattern analysis, dependency modeling, prediction, temporal DM, text mining

Recommendations for students	Association rules, clustering, sequential pattern analysis
Predicting student's performance	Neural networks, Bayesian networks, rule-based systems, regression, correlation analysis
Student Modeling	Naïve Bayes, Bayes net, SVM, logistic regression, decision trees
Detecting undesirable student behaviors	Classification, clustering, association rule
Grouping students	Classification (supervised learning), clustering (unsupervised learning)
Social networks analysis	Collaborative filtering
Developing concept maps	Association rules, text mining
Constructing Courseware	Clustering, classification, collaborative filtering
Planning and scheduling	Association rules

In the task 'Grouping Students' they mention different techniques to cluster/group students according to different purposes. The EM (Expectation-Maximization) algorithm mentioned in their review of the state of the art was used in this work [Myller, Suhonen and Sutinen 2002].

Providing that there is a mobilization to promote the massive use of discussion forums, it is a fact the huge increase on the volume of messages to be read and checked by teachers and tutors. Thus, some works have proposed ways to help teachers to keep up with students' messages using data mining techniques and text mining algorithms. Rodrigues and Oliveira (2014) described a system composed by several modules that aimed at creating and assessing exams as well as monitoring student's progress.

Furthermore, there are works which cluster the messages by genre [Guimarães, Esmin 2014] or by message types [Rolim, Cordeiro and Ferreira 2014], or even by graphs which show the relevance of what was written in the forum in order to engage students in a debate and to focus in concepts that the teacher considers important [Azevedo, Behar and Reategui 2011]. In their work, Machado *et. al* (2016) also used clustering techniques to infer some behavior aspects of students from their posts in the forums.

There are also published papers about text mining that aimed at classifying educational texts [Webber, Cristofoli and Lima 2013], or classify student's messages [Bovo *et al.* 2013, Raghavan *et al.* 2010], clustering and monitoring discussion forums [Oliveira Júnior *et al.* 2011, Oliveira Júnior and Esmin 2012], enhancing feedbacks in written works [Villalón, Kearney, Calvo and Reimann 2008], and helping to predict students' performance [Romero *et al.* 2013], among others. Nevertheless, they do not provide recommendations or any other kind of help regarding students' grammatical mistakes.

2.2 The Grammar Checker CoGrOO

Considering the great amount of written texts coming from the discussion forums, there is an impression that Portuguese grammar mistakes could be easily found. However, owing to the fact that teachers have to deal with a great number of messages and that these messages are posted in a public environment, shared by teachers and virtual class mates, it is difficult and even embarrassing to correct all the little and serious grammar errors in the middle of a discussion.

Orthography and grammar checkers are types of applications that use Natural Language Processing (NLP), an area connected to the field of Artificial Intelligence

(AI). And, as everything in the area of AI, they try to reproduce human abilities of doing or solving things, particularly in this case, to process and generate natural language. This project used the open source tool CoGrOO that makes and hybrid analysis considering statistics techniques of NLP and a rule-based approach to identify grammar mistakes [Silva 2013]. The main mistakes identified by this tool are: pronoun placement, nominal agreement, verbal agreement, the use of the grave accent (crase), use of ‘a/ há’, among others.

The Grammar Checker CoGrOO was developed by a group of NLP researchers with the purpose of creating a Brazilian Portuguese grammar checker to accompany the OpenOffice.org and its successor, the LibreOffice. The CoGrOO can identify a list of twenty errors, as shown in Table 1, and their combinations. These errors categories are based on the PROBI corpus [Martins 2002], a corpus composed of both correct and incorrect sentences collected to evaluate ReGra, a proprietary Brazilian Portuguese grammar checker.

Table 2. The error categories of CoGrOO based on PROBI corpus.

Error Number	Category	Description	Error Number	Category	Description
1	Adv	<i>Use of adverbs</i>	11	cov reg	<i>Verb Fazer</i>
2	adv con	<i>Agreement of adjective and noun</i>	12	cra	<i>Use of crase</i>
3	Aha	<i>Use of há/a</i>	13	ger	<i>Use of gerund</i>
4	Ali	<i>Others</i>	14	mal	<i>Use of mal/mau</i>
5	Cjc	<i>Use of conjunctions</i>	15	pro	<i>Use of pronouns</i>
6	Cmt	<i>Agreement between verbal modes and tenses</i>	16	ptn	<i>Use of punctuation</i>
7	Con	<i>Nominal agreement</i>	17	reg	<i>Verbal regency</i>
8	Cop	<i>Pronoun Placement</i>	18	ren	<i>Nominal regency</i>
9	cop pro	<i>Use of mim and ti</i>	19	sem	<i>Severe pleonasm</i>
10	Cov	<i>Verbal agreement</i>	20	ver	<i>Use of verbs</i>

The tool takes the user’s text as an input and gives, as an output, a list of possible mistakes. To do so, it executes a shallow parsing and then a rule-based checking. At first, it goes through a text analyzer with NLP techniques. Then, the text goes through a pipeline of annotators to check sentences tokens and boundaries, to attribute a POS (Part-of-speech class) tag for each word, and to find phrase chunks and subject-verb interrelations. After that, it matches a set of man-made errors rules with the resultant structure [Silva 2013].

This project used the CoGrOO API. This API integrated the solution proposed and it was responsible for finding the mistakes in the texts posted by the students in the forums of the LMS environment, in this case, Moodle. From these mistakes, data mining techniques were used to cluster students according to their difficulties and a specific support will be sent to each group in order to help them to overcome their grammatical problems.

3. The Proposed Workflow

This section presents the workflow of the solution proposed. For this work, a historical database was considered to find the clusters.

As it is shown in Figure 1, data is collected from the Moodle database. Some tables are selected through SQL scripts. These tables go through a preprocessing stage where attributes concerning students' personal information and their posts in the discussion forums are selected. A txt file is saved containing all the texts posted by a student in the discussion forum. The txt file receives the name of the student plus his/her e-mail. This information will be retrieved afterwards to send recommendations to each student according to the cluster they are in.

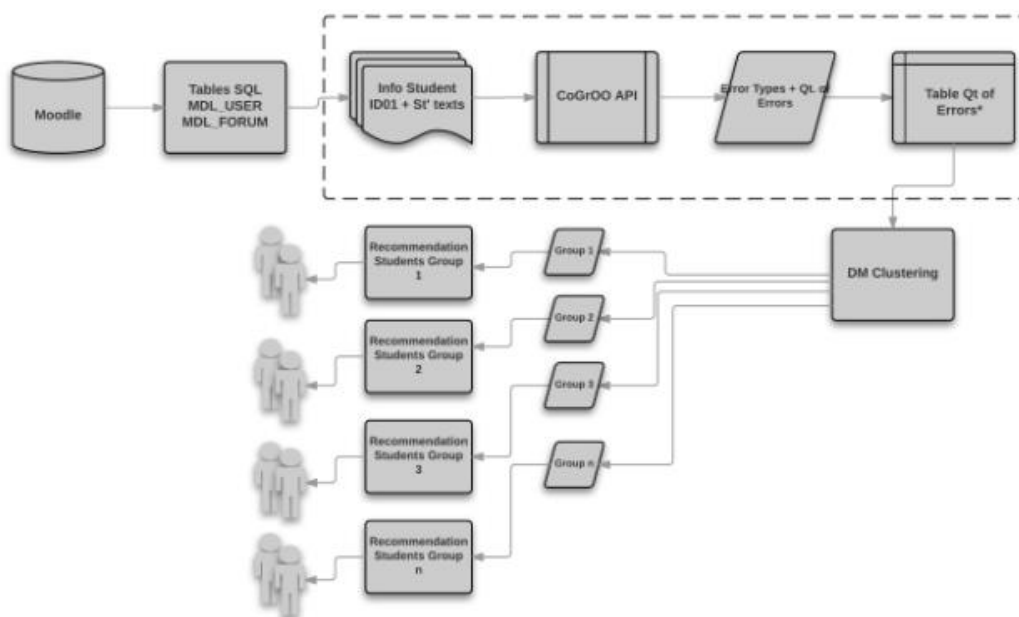


Figure 1. The project workflow

A Java application makes this process repeats for every student in the database. After having all the txt files, another Java application accesses the CoGrOO API and passes all the text files through it. The CoGrOO API delivers the list of mistakes for each txt file. This information is saved in a table.

The rows of this table list the students. The columns list the number of occurrences of each mistake. This step is over when all the txt files pass through the grammar checker. Then, this table goes through a preprocessing stage to prepare it for the clustering algorithm. The number of mistakes is normalized between 0 and 1. For this work, the EM (Expectation- Maximization) clustering algorithm was used. After having the students in clusters, specific recommendation will be sent to help students overcome their grammar problems. This paper covers this workflow up to the point where the clusters are formed.

4. Experiment

A historical database from the Núcleo de Educação a Distância (NEAD) of the University of Pernambuco was used for the experiment. This section will describe the process from selecting the data in the Moodle database to the use of the EM clustering algorithm to group students according to the mistakes they make and it is described as follows: the first part explains how the data was extracted and transformed from the

Moodle database; the second section shows how the CoGrOO API was integrated and how the students' mistakes were catalogued in a table; and the third section explains how the clustering algorithm was used in this project and the outputs are analyzed.

4.1 Database Preparation and Transformation

This data was extracted by using SQL scripts from a Moodle (version 2.7) database. The course of Pedagogy was selected to be analyzed. Figure 2 shows the steps of the data extraction. Firstly, all the information from all the semesters was listed. Then, one of the semesters was chosen to be analyzed. The 6th semester was chosen because it had a good number of disciplines, so it would probably have a good numbers of posts to go through de grammar checker. With the 6th semester ID, it was possible to list all the disciplines from this semester, including their IDs and names.

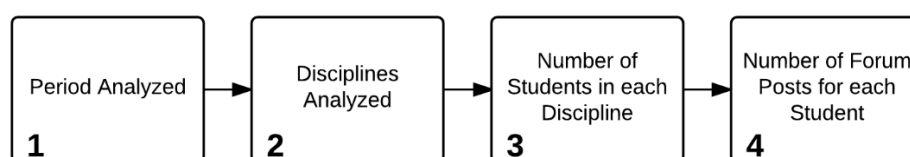


Figure 2. The steps of the data extraction

The next step was to list the number of students in each discipline. A Java script was created to make this process automatic. So, the program would look for all the students in each discipline. Another script searched for all the posts of a student. A txt file was created and named after each student. All his/her posts were put there. As the script was running at the posts' level, it was not a problem if a student belonged to more than one discipline, because all his/her posts from all the disciplines he/she was enrolled would be found and saved. It is important to mention that 9 disciplines were found in the 6th period, but only in 8 there were students enrolled, so the one without students was discarded. The total number of posts extracted was 9,323 for 357 students.

After this step, it was necessary to analyze the data to transform them. During these analyses it was observed that 91 students did not post anything in the forums. So, these students would not have any posts to go through the grammar checker and thus, they were taken out of the analysis. Then, 13 tutors and teachers were identified among the students due to their behavior and comments in the posts. They were also taken out from the data to be analyzed by the CoGrOO API. Finally, an outlier was found and also discarded. After this process, 252 txt files were ready to go through the grammar checker CoGrOO API.

4.2 The Grammar Checker

A Java script was created to make the 252 txt files go through the CoGrOO grammar checker. Firstly, to use the CoGrOO API it was necessary to compile it and to do so the Apache Maven was used. Then, it was necessary to import the CoGrOO API and the Apache POI API to the project.

The grammar analyzer was created using the CoGrOO. It was responsible for analyzing the texts and returning the grammar mistakes that were found. For each student there was a txt file with all the texts he/she had published in the discussion forums. This text was broken into paragraphs and the paragraphs went through the

grammar checker. For each paragraph analyzed, the grammar checker returned a list of mistakes. The list of mistake of each paragraph was added to the student's list of mistakes.

There were twenty types of errors distributed in 129 XML rules. Having all texts analyzed, the mistakes found were distributed in those 20 types according to the XML rule. After this process, an XLS file was created listing all the students in the rows and all the errors they made distributed in those 20 types in the columns.

From the 252 txt files that went through the CoGrOO, only two returned a problem. So, 99.2% of the data were processed by the grammar checker. The 0.8% that were not processed do not interfere in the result of the experiment because it is an insignificant amount and thus, analyses to understand why these two files did not pass through the CoGrOO were left to be done in the future.

4.3 The EM Algorithm, Results and Discussions

The XLS file created with the output data from CoGrOO had to be preprocessed and normalized to be ready for the clustering algorithm. Unnecessary information for this stage, such as the columns with the students' names and emails were taken out of the table. Also, some columns where no mistakes were counted were removed as they were not relevant for the mining process. The columns removed represented the errors numbers 2, 4, 11, 12, 15, 18 and 20. So, the data mining process considered only the mistakes numbers 1, 3, 5, 6, 7, 8, 9, 10, 13, 14, 16, 17, 19, totalizing 13 attributes and 250 instances (the total number of files/students checked by CoGrOO) (see table 2 to identify the errors).

A linear normalization process was done and the normalized data was imported to WEKA 3.8 for statistics analyses. The EM algorithm selected 3 clusters by cross validation, performing one iteration. The first cluster represented 53% of the instances (or students), with 132 instances. The second cluster represented 41%, with 102 instances, and the third cluster represented 6%, with 16 instances. Figure 3 shows the clusters.

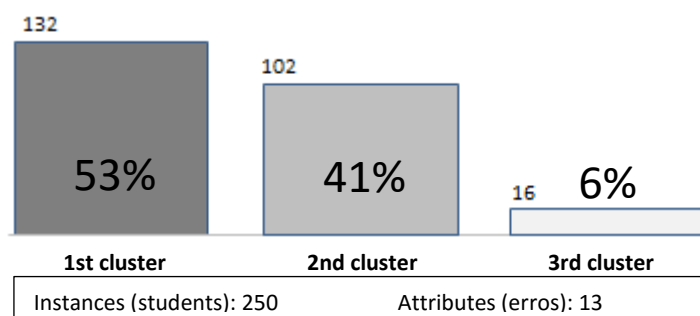


Figure 3. The output of the EM.

For the first cluster it was possible to observe that no students made the errors 5 (use of conjunctions), 6 (agreement between verbal modes and tenses) and 19 (severe pleonasm). For the errors 1 (use of adverbs), 13 (use of gerund), 16 (use of punctuation) and 17 (verbal regency), there was a good number of students who did not make or rarely made these mistakes with a mean of 0.52 errors by student. For the errors 3 (Use of *há/a*), 9 (Use of *mim* and *ti*) and 14 (use of *mal/mau*), almost any student made them

with a mean of 0.01 errors by student. The errors 7 (nominal agreement) and 10 (verbal agreement) were the ones with the highest means, 20.15 and 6.05 mistakes by students, respectively.

In 5 attributes of the second cluster no students made mistakes (errors: 1 (use of adverbs), 3 (Use of *há/a*), 6 (agreement between verbal modes and tenses), 9 (use of *mim* and *ti*), and 19 (severe pleonasm)). In others 6 attributes the mean of the mistakes by student was very low, 0.07 (errors: 5 (use of conjunctions), 8 (pronoun placement), 13 (use of gerund), 14 (use of *mal/mau*), 16 (use of punctuation), and 17 (verbal regency)). The only meaningful errors were the 7 (with a mean of 8.5 by student) and the 10 (with a mean of 2.62 by student). Therefore, this cluster could be considered the one with the students who made the least mistakes.

The third cluster, despite of being the smaller one in number of instances, it had the highest means for the errors 7 (nominal agreement) and 10 (verbal agreement), 20.56 and 6.31 respectively. Only for two attributes, the errors 5 (use of conjunctions) and 14 (use of *mal/mau*), there were no mistakes. There were occurrences for all the other attributes. The Table 3 shows the number of occurrences and the means for the two most significant errors: the error 7 and the error 10.

Table 3. The two most significant errors (occurrences and means)

	Number of Occurrences Error 7	Mean Error 7	Number of Occurrences Error 10	Mean Error 10
1 st Cluster	2,660	20.15	799	6.05
2 nd Cluster	867	8.50	268	2.62
3 rd Cluster	329	20.56	101	6.31

According to the results analyzed, all the students showed difficulties in nominal agreement and verbal agreement. Then, recommendations of studies regarding these grammar topics should be considered for all clusters. For the first cluster, the recommendations should focus on difficulties regarding the uses of gerunds and adverbs according to the means of errors in this cluster. For the second cluster the focus should be on pronoun placement, the most relevant error after the errors 7 and 10. For the third group, students should learn more about pronoun placement and verbal regency so, help on these matters should be considered.

5. Conclusions

Data mining of educational issues has been increasingly used to help actors in Distance Learning to better profit of this modality of teaching and learning. This paper aimed at presenting a workflow where open source applications could be integrated to offer help to students who have difficulties with the Brazilian Portuguese grammar. It related EDM techniques with a grammar checker to gather students according to their mistakes through the Expectation-Maximization algorithm.

The results showed that with the output of the grammar checker it is possible to apply mining techniques to extract knowledge from what students write on the discussion forums regarding grammar mistakes. Clustering students helped not only to identify the main difficulties of the whole sample, but also to find specific problems in

each cluster. As a future work, more detailed analyses will be done to find the correlations between the errors in each one of the clusters.

In the next phase of the project, the solution will be applied to a real environment, when the students will be able to receive the recommendations. Follow up analyses will be conducted in order to measure the effects of the recommendations received by the students.

Acknowledgements

This work was supported by NEAD (Núcleo de Educação a Distância), the Distance Learning Department of the University of Pernambuco.

References

- ABED – Associação Brasileira de Educação a Distância. Censo EaD.br: relatório analítico da aprendizagem a distância no Brasil 2013. Curitiba: Ibpex, 2014.
- AMARAL, Rogério do. As deficiências na produção do texto escrito: um estudo de caso referente aos graduandos de comunicação social. *Identidade Científica.V.2. N.2* , p.221-235. Presidente Prudente-SP: 2011.
- AZEVEDO, Breno F. T; BEHAR, Patrícia A.; REATEGUI, Eliseo B. Análise das mensagens de fóruns de discussão através de um software para mineração de texto. In: Anais do Simpósio Brasileiro de Informática na Educação - XXII SBIE – XVII WIE. Aracaju: 2011 .
- BOVO, Angela; SANCHEZ, Stephane; HÉGUY, Olivier; DUTHEN, Yves. Clustering Moodle data as a tool for profiling students. 2nd International Conference on e-Learning and e-Technologies in Education. ICEEE – 2013. p. 121-126. Lodz, Poland: 2013.
- GUIMARAES, F. R. N. ; ESMIN, A. A. A. . Identificação Automática de Gêneros das Mensagens em Fóruns de Discussões do AVA. In: Anais do The Brazilian Conference on Intelligent Systems (BRACIS) and Encontro Nacional de Inteligência Artificial e Computacional. São Carlos-SP: 2014.
- INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Censo da educação superior 2012: resumo técnico. Brasília: Inep, 2014.
- LIN, Fu-Ren; HSIEH, Lu-Shih; CHUANG, Fu-Tai. Discovering genres of online discussion threads via text mining. *Computers&Education*. 52(2009). p. 481-495.Elsevier: 2009.
- MACHADO, Crystiano J.R.; LIMA, Bruno R. B.; MACIEL, Alexandre M. A.; RODRIGUES, Rodrigo L. An investigation of students' behavior in discussion forums using Educational Data Mining. The 28th International Conference on Software Engineering & Knowledge Engineering SEKE-2016. Califórnia-USA: Jul/2016.
- MARTINS, Ronaldo. Probi: um corpus de teste para o revisor gramatical regra. Disponível em: <http://nilc.icmc.usp.br/nilc/download/NILC-TR-02-10.zip>. Acesso em: 04/05/2016.
- MOTTA, Íride Luiza de O. M. Dificuldades na escrita dos alunos de ensino superior: uma análise das narrativas escritas dos alunos da Faculdade Eduvale. *Revista*

Científica Eletrônica de Ciências Sociais Aplicadas da Eduvale. Ano III. N. 5
Out/2010. ISSN1806-6283. Jaciara-MT: 2010.

MYLLER, Niko; SUHONEN, Jarkko; SUTINEN, Erkki. Using data mining for
improving web-based course design. In: Proceedings of the Int. Conference on
Computers in Education. p.959-964. Washington-DC: 2002.

OLIVEIRA JÚNIOR, Roberto L. de; ESMIN, Ahmed A. A. Monitoramento
Automático de Mensagens de Fóruns de Discussão Usando Técnica de Classificação
de Texto. In: Anais do Simpósio Brasileiro de Informática na Educação - XXIII
SBIE – XVIII WIE. Rio de Janeiro: 2012.

OLIVEIRA JÚNIOR, Roberto L. de; ESMIN, Ahmed A. A.; COELHO, Tiago A.;
ARAÚJO, Deivison L.; SILVA, Leandro A.; GIROTO, Rafaela. Uma ferramenta de
monitoramento automático de mensagens de fóruns em ambientes virtuais de
aprendizagem. In: Anais do Simpósio Brasileiro de Informática na Educação - XXII
SBIE – XVII WIE. Aracaju: 2011.

RAGHAVAN, Preethi, CATHERINE, Rose; IKBAL, Shajith; KAMBHATLA, Nanda;
MAJUMDAR, Debapriyo. Extracting Problem and Resolution Information from
Online Discussions Forums. 16th International Conference on Management of Data –
COMAD 2010. Nagpur, India: Dec, 2010.

RODRIGUES, Fátima; OLIVEIRA, Paulo. A system for formative assessment and
monitoring of students' progress. Computers&Education. 76(2014). p. 30-41.
Elsevier: 2014.

ROMERO, Cristóbal; LÓPEZ, Manuel-Ignacio; LUNA, Jose-María; VENTURA,
Sebastián. Predicting students' final performance from participation in on-line
discussion forums. Computers&Education. 68 (2013). p. 458-472. Elsevier: 2013.

ROMERO, Cristóbal; VENTURA, Sebastián. Educational Data Mining: A Review of
the State of the Art. IEEE Transactions on Systems, Man and Cybernetics. Part C:
Applications and Reviews, Vol. 40, N.6, Nov/2010.

ROLIM, Vitor; CORDEIRO, Felipe. R.; FERREIRA, Rafael. Reconhecimento de
Padrões Aplicados a Comentários de Fóruns Educacionais. In: Encontro Nacional de
Inteligência Artificial e Computacional (ENIAC), 2014, São Carlos, SP. ENIAC
2014, 2014.

SILVA, William D. C. de M. Aprimorando o corretor gramatical CoGrOO. 2013. 166f.
Dissertação (Mestrado em Ciência da Computação). Instituto de Matemática e
Estatística, Universidade de São Paulo. São Paulo: 2013.

VILLALÓN, Jorge; KEARNEY, Paul; CALVO, Rafael; REIMANN, Peter. Glosser:
Enhanced Feedback for Students Writing Tasks. 8th IEEE International Conference
on Advanced Learning Technologies. Santander, Espanha: 2008.

WEBBER, Carine G.; CRISTOFOLI, Lauren G.; LIMA, Maria de Fátima W. do P.
UnderMine Text Miner – Uma ferramenta de mineração de texto para área
educacional. Novas Tecnologias na Educação. CINTED-UFRGS. Vol 11. N. 1. Rio
Grande do Sul: Jul/2013.